



TITLE:

# Local Economy and a Generalized Theory of Attract

AUTHOR(S):

Fujita, Koji

---

CITATION:

Fujita, Koji. Local Economy and a Generalized Theory of Attract. 英文学  
評論 1997, 69: 97-145

ISSUE DATE:

1997-03

URL:

[https://doi.org/10.14989/RevEL\\_69\\_97](https://doi.org/10.14989/RevEL_69_97)

RIGHT:

# Local Economy and a Generalized Theory of *Attract*

Koji Fujita

In this article, I further revise and improve some aspects of the recent development of the Minimalist Program advanced by Chomsky (1995). Major proposals include: (a) that both *Merge* and *Move* are subsumed under the general operation of *Attract*, and (b) that strictly local determination of optimal derivations can be implemented by the interaction of a small set of economy principles – the MLC, the MCC, and the MWC. These proposals constitute a further endorsement of the Minimalist claim that language is perfect.

## 1 Introduction

The Minimalist Program, as conceived by Chomsky (1993) and further developed by Chomsky (1994, 1995), takes the computational system of the human language faculty ( $C_{HL}$ ) as an internally coherent optimal system with its properties determined by requirements from other systems outside, i.e., by bare output conditions (BOCs). In particular, Chomsky (1995) argues that  $C_{HL}$  is uniform and inclusive in the derivation from the initial numeration  $N$  to the representation  $\lambda$  at the conceptual-intentional interface level of LF;  $C_{HL}$  does nothing but to arrange the lexical items selected for  $N$  to form  $\lambda$

that satisfies Full Interpretation (FI). Furthermore, of all such convergent derivations,  $C_{HL}$  picks out only the most economical one as admissible, to be associated with a (well-formed) linguistic expression. In this sense, each linguistic expression is an optimal realization of interface conditions. The study of the human language faculty is therefore largely concerned with the elucidation of the nature of the derivational process employed by  $C_{HL}$  and of the economy principles whose interactions serve to determine the optimality of a given derivation.

In this article, I will address these issues mainly by further revising Chomsky's (1995) new system and argue, in particular, (a) that *Attract* is the only kind of derivational operation available in  $C_{HL}$ , integrating *Merge* and *Move*, and (b) that maximally simple determination of an optimal derivation becomes possible under the interaction of certain local economy principles to be discussed — the Minimal Link Condition (MLC), the Maximal Checking Condition (MCC), and the Minimal Weight Condition (MWC). The approach to be proposed dispenses with the notion of global economy altogether, notably the commonly assumed requirement that derivations converge in the fewest steps possible, and limits necessary economy comparisons to strictly local ones. It also renders unnecessary some auxiliary assumptions adopted by Chomsky (1995) in order to implement the right kind of economy calculations under his system, including the claim that feature mismatch cancels the derivation. Thus our approach will constitute a viable solution to the problem of computational complexities that has come into the spotlight in recent literature (Chomsky 1996, Collins 1996, etc.). The general picture which will emerge as we proceed is this:  $C_{HL}$  has a unique derivational operation, and an optimal derivation results only from applying the most economical step at every stage. After all, this is what an optimal theory of

CHL should look like, in conformity with the Minimalist view that language is ‘almost’ perfect (Chomsky 1995).

The rest of the paper is organized as follows. Section 2 roughly sketches the view of CHL entertained by the current Minimalist theory (Chomsky 1993, 1994, 1995), and calls into question Chomsky’s (1995) distinction between Merge and Move, together with his reformulation of the MLC as part of the definition of Attract/Move. Section 3 proposes to integrate Merge and Move under the generalized operation of Attract. This generalized theory of Attract maintains that every application of Merge and Move is a last resort to let the derivation converge, and is therefore subject to the same set of economy principles. An optimal derivation will result from a sequence of the most economical step at each stage. Section 4 illustrates how this kind of strictly local economy comparison can be implemented under the interaction of the MLC, the MCC and the MWC. In particular, it will be argued that every step must satisfy all of these principles unless otherwise the derivation crashes. Section 5 extends this line of analysis to exclude movement into a  $\theta$ -position. Section 6 confirms the conceptual (and empirical) advantage of our approach by showing that it explains in a unified manner why a transitive construction never seems to allow the pair of accusative Subj-nominative Obj. Section 7 demonstrates that the MLC, as an economy principle, can be duly overridden by convergence. Section 8 concludes the paper.

## 2 Merge, Move, and Attract

Chomsky (1994) assumes *Merge* and *Move* as two different operations in CHL. Suppose  $\alpha$  and  $\beta$  are either a lexical item in the numeration  $N$  or an already constructed phrase marker  $\Sigma$ . Merge applies to  $(\alpha, \beta)$  and form a new

phrase marker  $\Sigma' = \{ \gamma, \{ \alpha, \beta \} \}$ , where  $\gamma$  is the label of  $\Sigma'$  and  $\gamma$  is either  $H(\alpha)$  (head of  $\alpha$ ) or  $H(\beta)$ . When  $\alpha$  and  $\beta$  are both terms of the same phrase marker  $\Sigma$ , the operation is called Move. Chomsky (1994) shows that some fundamental properties of Move derive from independent considerations. Among others, when  $\alpha$  moves by targeting  $\beta$  and forming  $\Sigma'$ , it must be the case (a) that  $\beta = \Sigma$  (Strict Cycle or the Extension Requirement; Chomsky 1993), and (b) that  $\gamma = H(\beta)$  (obligatory projection of the target). Chomsky (1993, 1994) proposes that Merge is costless, whereas Move is a costly operation subject to economy principles including Greed, Procrastinate, and the MLC.

Chomsky (1995) conceives the operation of Move in a rather different way. He first notes that, in principle,  $\alpha$  and  $\beta$  can also be a formal feature of a lexical item (F(LI)) or a bundle of such features (FF(LI)). Then he claims that Move can be better formulated in the general form of Attract, according to which the effect of  $\alpha$  raising to target  $\beta$  is a result of a sublabel of  $\beta$  attracting  $F(\alpha)$  or  $FF(\alpha)$ . Convergence at the articulatory-perceptual interface level of PF requires that Attract usually manifest itself in the form of a category movement when it applies before Spell-Out. Importantly, both Greed and the MLC are abandoned as economy principles and are incorporated into the definition of Attract/Move under this new system. Chomsky (1995) states:

- (1) K attracts F if F is the closest feature that can enter into a checking relation with a sublabel of K. (p.297)
- (2) K attracts  $\alpha$  only if there is no  $\beta$ ,  $\beta$  closer to K than  $\alpha$ , such that K attracts  $\beta$ . (p.311)

This Attract-F analysis is a further confirmation of the internal coherence of

CHL, for it reduces the varieties of movement operations to a single general form, with bare output conditions explaining the existence of such superficial differences, both among and within languages.

Furthermore, the substantial effect of Procrastinate is now captured by the general economy considerations concerning the 'weight' of what is being attracted. Watanabe (1993) has already proposed Economy of Weight (3) as an economy principle, in part to force excorporation where possible under his analysis of head movement, the details of which do not need to concern us here.

### (3) Economy of Weight (EOW):

Movement of heavier material is more costly. (Watanabe 1993 ; 161)

As originally formulated, EOW applies to compare movements of different sets of lexical categories, choosing movement of the fewest possible lexical items as optimal. Under the Attract-F analysis, EOW is naturally extended to comparisons of movements of different sets of features by stating that attracting more features is more costly. From this economy principle, to be dubbed the Minimal Weight Condition below, is derived a strictly local version of Procrastinate, just as the original Procrastinate was arguably a consequence of EOW, where PF convergence forces the more costly category movement over feature raising. I will return to this point below.

Attract-F guarantees that CHL is uniform from N to  $\lambda$  both pre- and post-Spell-Out, with the two derivational operations of Merge and Attract/Move. Maximal uniformity of the system would require further simplification, however, tempting us to unify these two operations. In this section, I will address this issue first by claiming that Chomsky's (1994, 1995) objection to such unification is not well-grounded. I will show that his

crucial data can be better accounted for if we assume Attract as the only operation in  $C_{HL}$ . His reformulation of the MLC as part of the definition of Attract/Move (see (1) and (2)) will also be called into question, which will be more fully examined in section 3.

## 2.1 Merge and Move: Two Different Operations?

For Chomsky (1994, 1995), there are both conceptual and empirical reasons to distinguish between Merge and Move. On one hand, Merge is primarily motivated by  $\theta$ -marking requirements or, more generally, by 'integration' in the sense of Collins (1995) and not by any feature checking requirement. Thus under the  $v$  (small V) analysis of transitive (and unergative) verbs adopted by Chomsky (1995), both merging Obj with V and merging Subj with  $v$  P are required for the purpose of proper  $\theta$ -marking (4a), whereas object shift to (the outer) [Spec,  $v$  P] is solely motivated by the checking of the strong D-feature of  $v$  (4b).

- (4) a. [ $v$  P Subj [ $v'$   $v$  [VP V Obj ]]]  
 b. [ $v$  P Obj [ $v'$  Subj [ $v'$   $v$  [VP V t<sub>Obj</sub> ]]]]

This complementarity between feature checking and  $\theta$ -marking, whether correct or not as such, does not suffice to discriminate the two operations. There are well-known instances of Merge required solely by feature checking, including merger of a pure expletive.

- (5) a. [TP There T is someone in the room ].  
 b. [TP<sub>1</sub> There T<sub>1</sub> seems [TP<sub>2</sub> t T<sub>2</sub> to be someone in the room ]].

In (5a), *there* is directly merged to [Spec, TP] to check the strong D-feature (the EPP-feature) of T, while in (5b), it first checks the EPP-feature of T<sub>2</sub> by

Merge and then moves to check that of T1. Merge in these examples is motivated by feature checking, on a par with Move in general, and the first obstacle to the integration of Merge and Move is removed. Rather, it seems fair to say that both Merge and Move are operations equally motivated by convergence.<sup>1</sup>

## 2.2 Is Merge Costless?

On a more empirical side, Chomsky (1994, 1995) argues that Merge must be kept separately from Move because only the former is costless and not subject to Procrastinate. This claim is based crucially on his proposed account of the following contrast involving an expletive.

- (6) a. [TP1 There T1 seems [TP2 *t*<sub>there</sub> T2 to be someone in the room ]].  
       cf. There seems to be someone in the room.  
       b. \* [TP1 There T1 seems [TP2 someone T2 to be *t*<sub>someone</sub> in the room ]].  
       cf. \*There seems someone to be in the room.

Chomsky (1994, 1995) proposes to block the derivation of (6b) in terms of a local economy comparison on grounds that it applies the more costly operation of Move to check the EPP-feature of T2, whereas merger of the expletive in (6a) is costless. In other words, (6b) violates Procrastinate in the lower clause. I will argue against this account in favor of the generalized Attract analysis to be proposed, according to which (6a,b) equally involve an application of Attract to fill [Spec,TP2] and any economy difference between them should be reduced to the weight of what is being attracted.

To see the problem with Chomsky's account of (6a,b), however, we first need to consider another type of contrast that Chomsky (1994, 1995) discusses in detail, which involves Superraising/Superpassive.<sup>2</sup>



- (7) a. [TP<sub>1</sub> It T<sub>1</sub> seems [CP [TP<sub>2</sub> John T<sub>2</sub> was arrested t<sub>John</sub> ]]].  
       cf. It seems John was arrested.
- b. \* [TP<sub>1</sub> It T<sub>1</sub> seems [CP [TP<sub>2</sub> t<sub>it</sub> T<sub>2</sub> was arrested John ]]].  
       cf. \*It seems was arrested John.
- c. \* [TP<sub>1</sub> John T<sub>1</sub> seems [CP [TP<sub>2</sub> it T<sub>2</sub> was arrested t<sub>John</sub> ]]].  
       cf. \*John seems it was arrested.

The grammaticality of (7a) poses an apparent puzzle for Chomsky, because its derivation involves the more costly operation of Move to check the EPP-feature of T<sub>2</sub> in the lower clause, exactly as in (6b); by contrast, (7b,c) result from applying the allegedly costless operation of Merge instead, as in (6a). Chomsky attempts to solve this near contradiction by noting that, of (7a-c), only (7a) represents a legitimate and convergent derivation. His argument goes as follows.

(7b,c) share the following structure before the creation of [Spec,TP<sub>1</sub>].

- (8) [TP<sub>1</sub> T<sub>1</sub> seems [CP [TP<sub>2</sub> it T<sub>2</sub> was arrested John ]]]

To check the EPP-feature of T<sub>1</sub>, either *it* or *John* has to move to fill its Spec, yielding (7b) or (7c). The MLC prohibits movement of *John* in favor of moving *it*, and therefore (7b) blocks (7c). The question is then why (7b) does not block (7a) as well. Crucially, Chomsky (1995) claims that (7b) does not converge, suggesting that the Case feature of either T<sub>1</sub> or *John* remains unchecked. (7b) is not a competing derivation at all, given that only convergent derivations compete for economy (Chomsky 1993, 1994), and it does not block (7a). It also follows that the MLC cannot be formulated as an economy principle, because it favors the nonconvergent derivation (7b) over the convergent (7c).<sup>3</sup> Chomsky (1995) concludes that the MLC should be

incorporated into the definition of Attract/Move. As a result, (7a) is now correctly predicted to be an optimal derivation.

Here I would like to argue, contra Chomsky, that both (7b) and (7c) represent convergent derivations. There is every reason to believe that in (7b), exactly as in (7c), all formal features that are [-Interpretable] are properly checked. Chomsky's observation that the Case feature of either T1 or *John* is not checked cannot be correct, because nothing precludes the possibility that the formal features FF(*John*) raise to T1 to check the Case feature after Spell-Out. This covert checking of the Case feature of T is exactly what Chomsky (1995) proposes in his treatment of the expletive construction as in (5a,b). The assumption is that the pure expletive *there* has only the categorial D-feature, so that it checks only the EPP-feature of T; after all, this is all that is required before Spell-Out for convergence. Then after Spell-Out, the formal features of the associate raise to check the rest of FF(T).

Turning back to (7b), we see that *it* does not check the Case feature of T1 because, by checking that of T2, it has already lost its Case feature in the lower clause. The Case feature of T1 remains unchecked before Spell-Out, but this is fine as it is in (5). Then after Spell-Out, FF(*John*) raise and check the Case and other features of T1. (7b) and (7c) do converge, then, and this conclusion ruins Chomsky's account of the grammaticality of (7a) in a devastating manner. In particular, the reformulation of the MLC as part of the definition of Attract/Move is now without warrant. Rather, it must be the case that, of the three convergent derivations (7a-c), economy considerations pick out (7a) as optimal, which casts serious doubt on the proposed distinction between Merge and Move.

Below I will show that by abandoning this distinction we can account for

the data in (7) in an elegant way, in terms of the conspiracy of the MLC as an economy principle and another principle that favors an operation that checks more formal features of the target, the Maximal Checking Condition (MCC). At the same time, this line of analysis entails that the contrast in (6) can no longer be explained in the manner reviewed above. I will claim that it is rather Watanabe's (1993) EOW (3) or the Minimal Weight Condition (MWC) to be introduced below that accounts for the contrast. In the next section, I will first outline a generalized theory of Attract and clarify its conceptual naturalness.

### 3 A Generalized Theory of Attract

I have pointed out above that Chomsky's (1994, 1995) distinction between Merge and Move, taking only the latter to be a costly operation, does not work properly in accounting for the contrast in (7). I now propose that the distinction should be removed by subsuming both Merge and Move under the general operation of Attract. The concept of Attract has been introduced by Chomsky (1995) to replace that of Move; it is no longer that  $\alpha$  moves to  $\beta$  to satisfy its morphological properties, as prescribed by Greed (Chomsky 1993, 1994), but rather that some formal feature associated with  $\beta$  (a sublabel of  $\beta$ ) requires checking so that it attracts a corresponding feature of  $\alpha$ . To the extent that attracting a single feature or a feature bundle is prohibited for reasons of (PF) convergence, overt application of Attract takes the form of category attraction; where  $\alpha$  and  $\beta$  are terms of the same phrase marker, it captures the effect of the classical Move  $\alpha$ .

#### 3.1 Merger of an Expletive as Attract

Consider now Merge, and how closely its effect can be assimilated to that of

Move just described. We have already seen that there exists at least one instance of Merge that can be readily characterized in terms of feature attraction; merger of an expletive to check the EPP-feature of T.<sup>4</sup>

- (9) a. [TP T XP ] →  
 b. [TP Exp T XP]

In this particular case, it is only natural to say that the strong D-feature of T 'attracts' the D-feature of the expletive, pied-piping the whole category for convergence. Merge here differs from Move only in that the target attracts the relevant feature/category from the numeration N, not from within the same phrase marker as when T attracts Subj.

- (10) a. [TP T [<sub>vP</sub> Subj *v* VP ]] →  
 b. [TP Subj T [<sub>vP</sub> t<sub>Subj</sub> *v* VP]]

Apart from this difference, Merge in (9), on a par with Move in (10), can be thought of as an instance of Attract, which, if so, should be subject to general economy considerations.

The supposition that merger of an expletive obeys economy principles offers an elegant account of why merging it first in a lower position and then moving it to [Spec,TP] is more costly than, and therefore is blocked by, merging it directly in [Spec,TP].

- (11) a. [TP Exp T ... [XP X ... ]]  
 b. \* [TP Exp ... [XP t<sub>Exp</sub> X ... ]]

Suppose [Spec,XP] need not be filled for convergence; perhaps X is an unaccusative verb. To the extent that we adopt Chomsky's (1994, 1995) characterization of Merge as a costless operation, we have to resort to an

economy principle of global nature to block derivation (11b) in favor of (11a), say the Shortest Derivation Condition, which may be formulated as in (12).

(12) Shortest Derivation Condition (SDC):

A derivation in fewer steps is more economical.

(11b) involves at least one more application of Move than (11a), and therefore is blocked by the SDC. Notice, however, that the global SDC compares competing derivations in terms of the total numbers of their operations, something that we hope to be able to reduce to a principle of local economy to avoid introducing high computational complexity.

By taking Merge also to be a costly operation, it now becomes possible to block (11b) in a strictly local manner. At the very point of the derivation where XP is formed, we can exclude (11b) as more costly than (11a) simply on grounds that it applies Merge where not required by convergence. That is, (11b) violates a strong, strictly local version of Procrastinate.

(13) Strong Procrastinate (SP):

Do not attract.

SP holds of each step of the derivation, requiring that an operation apply as late as possible, not in terms of the pre-/post-Spell-Out distinction but literally step-wise. Given that convergence overrides economy (Chomsky 1994, 1995), SP may be violated only if the derivation would not otherwise converge. We now have rather strong conceptual reason to believe that merger of an expletive constitutes a subtype of the general operation of Attract. I return to the nature of SP below, where I will propose an alternative account of the contrast in (6), on which Chomsky (1994, 1995) builds his argument that Merge is costless.<sup>5</sup>

### 3.2 Merger of an Argument as Attract

With this much in mind, consider other instances of Merge, among others, merger of an argument into its  $\theta$ -position.

$$(14) \quad [_v P \text{ Subj } [_v' v \text{ } [_{VP} V \text{ Obj } ]]] \quad (= (4))$$

Merge here is triggered by the  $\theta$ -marking properties of the verbs. On the assumption that a  $\theta$ -Criterion violation causes the derivation to crash (Chomsky 1994, 1995), the operation is required by convergence and there is no danger in saying that merger of an argument is a costly operation subject to economy principles. One might even proceed to assimilate the  $\theta$ -grid of a verb, for example, to a strong feature, in the sense that it has to be satisfied immediately once it is introduced into a phrase marker. Thus Chomsky (1995: 234) states, of the derivation  $D$  that forms  $\Sigma$  containing  $\alpha$  with a strong feature;

$$(15) \quad D \text{ is canceled if } \alpha \text{ is in a category not headed by } \alpha.$$

To the extent that a  $\theta$ -role has to be discharged within the projection of the assigning head, it must be 'strong' in the relevant sense. This provides a strong motivation for the predicate-internal subject hypothesis under the current framework. We do not want to say that merging Subj directly in [Spec,TP] is more economical than merging it first in [Spec, $v$ P], in terms of SP or even the SDC (see Collins 1995 for related discussion).

Merger of an argument is then also a costly operation, to be subsumed under Attract, which is permitted to apply only when required by convergence. Here it is the  $\theta$ -marking property of the head (whether or not we are ready to call it a formal feature; see the discussion in section 7) that attracts the argument to its  $\theta$ -marking domain. This immediately excludes a

derivation in which an argument moves to its  $\theta$ -position from somewhere else.

- (16) \*<sub>[<sub>v</sub> P Subj v [VP ... t<sub>Subj</sub> ...</sub>

Merging an argument in a position other than its  $\theta$ -position is already a (potential) violation of SP. Furthermore, to the extent that movement from one  $\theta$ -position to another yields a  $\theta$ -Criterion violation, we maintain the correct generalization that merger of an argument is only possible in its  $\theta$ -position.

In (14), Subj and Obj can be a category directly drawn from N (when they are lexical categories), or they can be an already constructed phrase marker  $\Sigma'$  (when they are phrasal). In terms of Attract, this amounts to saying that the target may attract  $\alpha$  when  $\alpha$  is in N or is a  $\Sigma'$ . Merger of an expletive is an instance of Attract of the former type; Attract of the latter type can be exemplified by satisfaction of selectional properties.

- (17) a. I wonder [<sub>CP</sub> what Mary saw t ].  
 b. \*I wonder [<sub>CP</sub> Mary saw what ].

The descriptive fact is that *wonder* requires a [<sub>+</sub>wh] CP complement once it is introduced into the phrase marker. In other words, *wonder* must be merged with such a complement, and this requirement is not satisfied in (17b). Such a selectional property is again susceptible to an analogy with a strong feature, in the sense that it has to be satisfied within the projection of the head.

- (18) [<sub>VP</sub> wonder[<sub>+</sub>wh] [<sub>CP</sub>[<sub>+</sub>wh] what C[<sub>+</sub>wh] [<sub>TP</sub> Mary saw t ]]]

Let us assume, as seems natural, that the [<sub>+</sub>wh] feature of *wonder* attracts an

already constructed  $\Sigma' = \text{CP}_{[+wh]}$  for convergence. Note that, although the overt wh-movement to [Spec,CP] has already eliminated the strong feature of the head C, its [+wh] feature (perhaps as a categoral feature) remains accessible to the computation (see Chomsky 1995; sec.4.5.4.).

By extension, merger of an adverb can also be thought of as an instance of Attract, if an adverb is not free to occur anywhere in the phrase structure but has its own specified position to be licensed by the relevant head (see Bowers 1993, among many others).

### 3.3 Generalized Attract

Attracting  $\alpha$  seems also possible when  $\alpha$  is contained within an independent phrase marker  $\Sigma'$ . Bobaljik (1995) argues that this particular case of Attract occurs in head movement. Head movement has been known for its immunity from the Extension Requirement (ER) of Chomsky (1993). Suppose V raises overtly to T in the language.

- (19) a. [TP T [VP ... V ...  
       b. [TP [T V-T][VP ... tv ...

In (19a), T is merged with VP to form TP, satisfying ER. In (19b), however, subsequent movement of V to T does not result in extending the whole phrase marker. Rather than exempting head movement from ER as such, Bobaljik (1995) shows that it in fact obeys ER by allowing the derivation to proceed as in (20).

- (20) a. [VP ... V ...  
       b. [T V-T] / [VP ... tv ... (Merger of T and V)  
       c. [TP [T V-T][VP ... tv ... (Merger of T and VP)



The crucial step is (20b), in which V-raising targets T before they appear in the same phrase marker. Head movement here extends the category T by adjunction, which in turn is merged with VP to form a larger phrase marker by the next step (20c). Recast in terms of Attract, T attracts V from an already constructed phrase marker  $\Sigma'$ .

Assuming this kind of Merge to be another viable option, we can state the generalized theory of Attract as follows.

(21) Generalized Attract:

Attract is the only kind of operation in  $C_{HL}$ .

(22) Subtypes of Attract:

Move: The target attracts  $\alpha$  within the same phrase marker  $\Sigma$ .

Merge: The target attracts  $\alpha$  from the numeration N or an independent phrase marker  $\Sigma'$ , or  $\alpha = \Sigma'$ .

In the following discussion, I will continue to use the terms Merge and Move only for expository purposes, while reserving Attract to refer to the generalized operation that integrates both of them. By this integration, Merge is fully expected to be as costly as Move; they are subject to exactly the same economy principles and can be applied only when required by convergence.

### 3.4 Merge and the MLC

At this point in the course of the discussion, it is helpful to think about the relation between Merge and the MLC, since under our theory Merge is also subject to economy principles, including the MLC. Interestingly enough, it can be easily shown that Merge trivially satisfies the MLC almost by definition.

First let us formulate the MLC as an economy principle as in (23), with the relevant notion of 'closeness' stated as in (24), basically following Chomsky (1995).

(23) Minimal Link Condition (MLC):

Attracting a closer element is more economical.

(24)  $\beta$  is closer to  $\tau$  than  $\alpha$  only if  $\beta$  c-commands  $\alpha$  and  $\beta$  is not in the same minimal domain as  $\tau$  or  $\alpha$ .

Given that the configurational relation of c-command holds only within the same phrase marker  $\Sigma$ ,  $\beta$  cannot be closer to  $\tau$  than  $\alpha$  unless  $\alpha$  and  $\beta$  are both contained in  $\Sigma$ . When at least either  $\alpha$  or  $\beta$  is still in the numeration, neither can c-command and therefore be closer than the other. It is obvious then that Merge in the narrower sense could not possibly violate the MLC.

Consider Merge in the wider sense of Bobaljik (1995) just mentioned. Suppose first that T attracts V from an independent phrase marker  $\Sigma'$ , which happens to contain more than one instance of V.

- (25) a. T / [VP<sub>1</sub> V<sub>1</sub> ... [VP<sub>2</sub> V<sub>2</sub> ...  
       b. [T V<sub>1</sub>-T ] / [VP<sub>1</sub> t<sub>V<sub>1</sub></sub> ... [VP<sub>2</sub> V<sub>2</sub> ...  
       c. \*[T V<sub>2</sub>-T ] / [VP<sub>1</sub> V<sub>1</sub> ... [VP<sub>2</sub> t<sub>V<sub>2</sub></sub> ...

In (25a),  $\Sigma' = \text{VP}_1$ , which contains V<sub>2</sub> in addition to its own head V<sub>1</sub>. To the extent that V<sub>1</sub> c-commands V<sub>2</sub>, attracting V<sub>2</sub> rather than V<sub>1</sub> violates the MLC, yielding the usual Head Movement Constraint effect. Suppose next that T in (25b) attracts some VP. One may fairly expect that it is again the MLC that chooses attraction of VP<sub>1</sub> over attraction of VP<sub>2</sub> as more economical.

- (26) a. [TP [T V<sub>1</sub>-T ] [VP<sub>1</sub> t<sub>V<sub>1</sub></sub> ... [VP<sub>2</sub> V<sub>2</sub> ...

b. \* [TP [T V1-T ] [VP2 V2 ... / [VP1 tv1 tvP2

If it is appropriate to say that in (25b) VP1 is closer to T than VP2 (though this requires partial modification of the definition of c-command), the derivation in (26b) is immediately blocked by the MLC. We now understand that the generalized Attract analysis allows us to reduce the root nature of Merge (Chomsky 1994, 1995) at least in part to the MLC.

In short, subjecting Merge to the MLC does not give rise to unnecessary computational burdens, as one might suspect. Rather, it promises to successfully derive some properties of Merge that have been taken to distinguish it from Move without any descriptive discrimination between the two. This much said, we are now in a position to see how neatly the generalized theory of Attract explains the contrasts in (6) and (7).

#### 4 Local Determination of Optimal Derivations

In this section, I will outline an account of (7a-c), and then of (6a,b), under the proposed generalized theory of Attract. Recall that Chomsky's distinction between Merge and Move, and also his reformulation of the MLC as part of the definition of Attract/Move, crucially depend on the correctness of his analysis of these data, which has been cast into serious doubt in the foregoing discussion.

In section 4.1, I will first suggest an account of (7a-c) in terms of the global economy principle SDC (12) as it interacts with the local principle MLC (23). Noting the conceptual inadequacy of the SDC, however, in section 4.2 I will introduce another local principle, the Maximal Checking Condition (MCC), to replace the SDC. This will enable us to restrict ourselves to strictly derivational economy comparisons in the determination of an optimal

derivation. Then in section 4.3, I will propose still another local principle, the Minimal Weight Condition (MWC), and argue that it is this principle that accounts for the contrast of (6a,b). Section 4.4 mentions a potential issue posed by positing both the MCC and the MWC, which will be addressed later in this paper.

It should be kept in mind that the general outlook of the theory pursued here goes as follows:  $C_{HL}$  contains only one operation of Attract (generalized Attract), integrating both Merge and Move, whose application is always subject to the same economy principles. An optimal derivation is one which takes the most economical step at every stage.

In the following discussion, I adopt Chomsky's (1995) elaborated theory of formal feature checking and assume that only [+Interpretable] features remain accessible to the computation even after being checked, whereas [-Interpretable] features are eliminated once checked. I also assume with Chomsky that feature checking takes place only under feature match. Later in section 6, however, I will depart from Chomsky (1995) and abandon (27).

(27) Mismatch of features cancels the derivation. (Chomsky 1995; 309)

Chomsky distinguishes between nonmatch and mismatch, with the former allowing the derivation to proceed. Under his system, cancellation amounts to convergence for economy comparisons in the sense that it blocks less economical derivations. On the other hand, I will argue that both the distinction between nonmatch and mismatch and the notion of cancellation under feature mismatch can be rejected in our approach.

#### **4.1 The SDC: (Failure of) Global Economy**

Consider first (7a-c), reproduced here as (28a-c) for convenience.

- (28) a. [TP<sub>1</sub> It T<sub>1</sub> seems [CP [TP<sub>2</sub> John T<sub>2</sub> was arrested t<sub>j</sub> ]]].  
 b. \*[TP<sub>1</sub> It T<sub>1</sub> seems [CP [TP<sub>2</sub> t<sub>it</sub> T<sub>2</sub> was arrested John ]]].  
 c. \*[TP<sub>1</sub> John T<sub>1</sub> seems [CP [TP<sub>2</sub> it T<sub>2</sub> was arrested t<sub>j</sub> ]]].

We have seen that, contrary to Chomsky's observation, (28a-c) all correspond to convergent derivations. The problem is therefore why (28a) can be optimal, despite the fact that this is the only derivation that has applied the allegedly more costly operation of Move in the lower clause. Under the proposed framework of generalized Attract, this problem does not arise in the first place, for Move in itself can no longer be more costly than Merge.

The relevant steps in the derivations of (28a-c) can be roughly represented as in (29)-(31), respectively.<sup>6</sup>

- (29) Derivation of (28a):  
 a. [TP<sub>2</sub> John T<sub>2</sub> was arrested t<sub>j</sub> ]  
 b. [TP<sub>1</sub> it T<sub>1</sub> seems [CP[TP<sub>2</sub> John T<sub>2</sub> was arrested t<sub>j</sub> ]]]
- (30) Derivation of (28b):  
 a. [TP<sub>2</sub> it T<sub>2</sub> was arrested John ]  
 b. [TP<sub>1</sub> it T<sub>1</sub> seems [CP[TP<sub>2</sub> t<sub>it</sub> T<sub>2</sub> was arrested John ]]]  
 (Spell-Out)  
 c. [TP<sub>1</sub> it FF(J)-T<sub>1</sub> seems [CP[TP<sub>2</sub> t<sub>it</sub> T<sub>2</sub> was arrested t<sub>FF(J)</sub> ]]]
- (31) Derivation of (28c):  
 a. [TP<sub>2</sub> it T<sub>2</sub> was arrested John ]  
 b. [TP<sub>1</sub> John T<sub>1</sub> seems [CP[TP<sub>2</sub> it T<sub>2</sub> was arrested t<sub>j</sub> ]]]

In (29), step (a) is a forced violation of Strong Procrastinate, to satisfy the EPP-feature of T<sub>2</sub>; the Case and  $\phi$ -features of T<sub>2</sub> are also checked at the

same time. Likewise, step (b) is required by convergence, resulting in the proper checking of all the formal features of T1. Note that both applications of Attract trivially satisfy the MLC. Ignoring all irrelevant steps, then, we could say that (29) is a convergent derivation with just two applications of Attract. In (30), step (b) does not suffice to check all the formal features of T1, because *it* does not have the nominative Case feature when it reaches [Spec,TP1]. The derivation proceeds, however, and T1 attracts FF(*John*) after Spell-Out, in particular its nominative Case, as in step (c). The derivation converges, then, but only with three relevant applications of Attract. At this point, one may be tempted to argue that (30) is blocked by (29) in terms of the SDC (12), which favors a derivation in fewer steps. Let us tentatively assume that this is so. Incidentally, no step in (30) violates the MLC, to the details of which I will return below. By contrast, (31) is a convergent derivation with two applications of Attract, and is as economical as (28) as far as the SDC is concerned. However, step (31b) is blocked by (30b) under the MLC.

In sum, (29) is more economical than (30) in terms of the global SDC, and (30) is in turn more economical than (31) in terms of the local MLC, which allows us to pick out (29) as optimal. Importantly, we cannot proceed to say that the MLC blocks (31) in favor of (29) by directly comparing them. The MLC, as a local principle, can only apply to a particular stage of a single derivation, determining the next step that is most economical. (29) and (31) are already two different derivations when step (31b) violates the MLC, however, and they cannot be subject to such a local comparison. Rather, it is only by comparison with step (30b) that the MLC blocks step (31b).

This last point is of much theoretical interest, because it is rather (31) that is more economical than (30) in terms of the SDC. Restricting our

attention to these two derivations, we seem to face a situation where the relative status of different economy principles is at issue. This topic deserves careful treatment because it directly bears on the nature of economy principles in the Minimalist framework.

Suppose, in general terms, that there are two economy principles,  $\gamma$ ,  $\delta$ , and two competing derivations,  $\Gamma$ ,  $\Delta$ , such that  $\Gamma$  is more economical under  $\gamma$  but  $\Delta$  is more economical under  $\delta$ . The relative status of  $\gamma$  and  $\delta$  is reflected by whether or not either derivation can be chosen as optimal. Adapting terminology from another recent approach to the topic of optimality in syntactic theory (Grimshaw 1993, Pesetsky 1994, etc.), we can state the conceivable ‘ranking’ relations as follows:

- (32) a.  $\gamma$  outranks  $\delta$  if only  $\Gamma$  can be optimal.  
 b.  $\gamma$  and  $\delta$  are negatively tied if neither  $\Gamma$  nor  $\Delta$  can be optimal.  
 c.  $\gamma$  and  $\delta$  are positively tied if both  $\Gamma$  and  $\Delta$  can be optimal.

From the viewpoint of minimum computational complexity, negative tying is the best case we should hope because it has the effect of precluding any derivation that is less economical in terms of some principle, whether or not it is more economical in other respects. In effect, it amounts to saying that an optimal derivation must be most economical with respect to every economy principle.

An examination of the brief discussion on (30) and (31) above reveals that the SDC does not outrank the MLC, because otherwise it would be impossible to choose (29) over (30) and (31) as optimal. Recall that the local nature of the MLC does not allow (29) to block (31), which therefore must be blocked by (30) under the MLC although the SDC chooses (31) over (30). For the same reason, the MLC and the SDC cannot be positively tied, either. We

understand then that the MLC and the SDC are negatively tied, or otherwise the MLC outranks the SDC.

There is conceptual reason to expect the former possibility to be true, but our analysis so far offers no empirical ground for this choice. I take this indeterminacy in the ranking relation between the MLC and the SDC, together with the global nature of the SDC itself, as decisive evidence against the SDC. After all, global economy is what we have to avoid to reduce computational complexity; it is not possible to determine whether or not a given derivation satisfies the SDC unless the total number of its steps can be compared with those of all other competing derivations, which is certainly not a preferable state of affairs. Although it seems possible to conjecture that the kind of parallel computation involved in calculating global economy may well be part of the human language faculty, I will abandon the SDC and replace it with another local principle – the Maximal Checking Condition (MCC).

## 4.2 The MCC: Local Economy

It is evident that attracting *it* to [Spec,TP1] in (30b) leads to a violation of the global SDC exactly because this operation is not sufficient to check all the formal features of the target. Specifically, the Case feature of T1 fails to be checked and, accordingly, an additional application of covert feature raising is required for convergence. Attracting *John* instead to [Spec,TP1], as in (31b), is sufficient in this respect; this step allows the checking of all the formal features of T1, and no subsequent operation is necessary. This observation makes it clear that the essential effect of the SDC is properly captured in local terms by the following economy principle:



## (33) Maximal Checking Condition (MCC):

Attracting an element that checks more formal features of the target is more economical.

The MCC applies to compare only steps (30b) and (31b) at the stage of the derivation where [Spec,TP1] is to be created, choosing the latter over the former. Because of this local nature, it is no longer possible either to compare derivations (29) and (30) or to block (30) in favor of (29). Rather, (30) and (31) block each other, as step (30b) is more economical than (31b) in terms of the MLC, but (31b) is more economical than (30b) in terms of the MCC.

The account just given elucidates the relative status of the MLC and the MCC; they are negatively tied in the sense clarified above. A derivation that violates either the MLC or the MCC at any stage cannot be picked out as optimal; the derivation is literally 'terminated' or 'deadlocked' at that stage, because it cannot proceed further by taking an optimal next step. This kind of local economy comparison is an outstanding merit of adopting the MCC instead of the SDC, and I take it to be strong support for the former.

In sum, our account of (28a-c), corresponding to (29-31), respectively, goes as follows. Step (29a) is as economical as (30a) and (31a), because Move is no longer more costly than Merge under the generalized theory of Attract adopted here, and also because it satisfies both the MCC and the MLC exactly as the other two do. These operations equally check all the formal features of the target T2, and *it* in the numeration is not closer to T2 than *John* in the already constructed phrase marker. Steps (29a), (30a) and (31a) are optimal in this sense, and all the derivations are allowed to proceed. But now derivation (29) departs from, and therefore can no longer be compared with,

(30) and (31). In other words, (29) is determined as an optimal derivation on its own, on grounds that every step it takes is the most economical one at each stage. In particular, step (29b) again satisfies both the MLC and the MCC for the now obvious reason. By contrast, neither (30b) nor (31b) is optimal because they violate the MCC and the MLC, respectively. The determination of an optimal derivation is simple enough here, with necessary economy comparisons restricted to strictly local ones, and (28a) is correctly predicted to be grammatical.

### 4.3 The MWC: Local Economy

Our analysis so far leaves unexplained the contrast of (6a,b), reproduced here.

- (34) a. [TP<sub>1</sub> There T<sub>1</sub> seems [TP<sub>2</sub> t<sub>there</sub> T<sub>2</sub> to be someone in the room ]].  
 b. \* [TP<sub>1</sub> There T<sub>1</sub> seems [TP<sub>2</sub> someone T<sub>2</sub> to be t<sub>someone</sub> in the room ]].

I now propose that it is basically Watanabe's (1993) EOW (3) that is responsible for this contrast. As noted above, under the current theory of formal feature attraction, EOW is naturally extended to compare attractions of different feature bundles, stating that attracting an element with more formal features is more costly. Let us formulate the relevant principle as follows:

- (35) Minimal Weight Condition (MWC):

Attracting fewer formal features is more economical.

The natural Minimalist assumption underlying the MWC is that C<sub>HL</sub> has access to, and therefore 'weighs,' only formal features.

Suppose at some stage of the derivation the target  $\tau$  need not attract

any formal feature for convergence. The MWC requires that  $\tau$  attract nothing; it prohibits any application of Attract. We now understand that Strong Procrastinate stated in (13) above is an automatic consequence of the MWC. Note that the MWC is also a strictly local principle, on a par with the MLC and the MCC.

Given the MWC, the explanation for the contrast of (34a,b) is straightforward. They share the following phrase marker  $\Sigma$  before [Spec,TP2] is created.

(36) [TP<sub>2</sub> T<sub>2</sub> to be [ someone in the room ]]

To check its EPP-feature, T<sub>2</sub> now attracts either *there* (from N) or *someone* (within  $\Sigma$ ). The former option is obviously more economical under the MWC, since the pure expletive *there* has only the categorial D-feature, but *someone* carries other features as well. Thus (34a) can be picked out as optimal by a simple local comparison without discriminating Merge and Move.

Chomsky (1995), in discussing the transitive expletive construction (TEC) in Icelandic, attempts to show why it always corresponds to the structure in (37) under his system.

(37) [TP Exp [T' Subj [T' T [XP t<sub>Subj</sub> X ... ]]]]

Here the expletive Exp and its associate Subj occupy the outer and inner [Spec,TP], respectively, and it is assumed that the N-feature of Subj raises to Exp after Spell-Out to check the [-Interpretable] D-feature of Exp; otherwise the derivation crashes. At the stage of the derivation where the inner [Spec,TP] in (37) is created, T may attract either Exp or Subj to check its EPP-feature; attracting Subj is more costly under the MWC, which might

appear to predict incorrectly that the structure of the TEC must be as in (38), instead:

(38) [TP Subj [T' Exp [T' T [XP t<sub>Subj</sub> X ... ]]]]

Derivation (38) does not converge, however, because here subsequent N-feature raising from Subj to Exp is inapplicable; lowering does not exist as a possible operation in the current framework. We understand then that the apparent violation of the MWC in (37) is a forced, and therefore permissible one, in accordance with the Minimalist claim that convergence overrides economy.

Now consider possible continuations of the derivation for (34a) after [Spec,TP2] is filled by the expletive. The relevant structure is (39), where T1 attracts some element to check its EPP-feature.

(39) [TP1 T1 seems [TP2 there T2 to be someone in the room ]]

The fact is that T1 must attract *there* rather than *someone*, which seems to pose a potential problem, given the conspiring relation of the MLC and the MCC we have exploited above. Notice that attracting *there* satisfies the MLC but violates the MCC, because attracting *someone* instead would allow all the formal features of the target to be checked at once. Why, then, do these two steps not block each other? The answer is already clear; in (39), attracting *someone* to [Spec,TP1] would cause the derivation to crash, because it would prohibit subsequent covert checking of the D-feature of *there*, just as in (38). In other words, the derivation which satisfies the MCC does not count as a competing derivation, and the MCC is rightly violated in (34a) for the purpose of convergence.

#### 4.4 The MCC vs. the MWC: A Conflict

Here a brief remark on the possible 'ranking' relations between the MCC and the MWC may be in order, because these two principles almost always contradict each other. Roughly speaking, the MCC favors attraction of more formal features, to check as many features as possible, whereas the MWC disfavors such an operation. These principles therefore cannot be negatively tied, because otherwise the derivation would never converge optimally where they conflict.

Suppose that a language has strong *v* and strong T, necessitating both overt object shift and the EPP effect. The said conflict occurs after object shift, when T is about to attract either Subj or Obj to its Spec:

- (40) a. [TP T [<sub>v</sub> P Subj<sub>[nom]</sub> *v* [VP V Obj<sub>[acc]</sub> ]]]  
       b. [TP T [<sub>v</sub> P Obj [<sub>v</sub>' Subj<sub>[nom]</sub> *v* [VP V t<sub>Obj</sub> ]]]

In (40b), Obj has checked the Case feature of *v*, among others, and lost its own Case feature consequently. It is therefore 'lighter' than the nominative Subj, and Obj raising to [Spec,TP] is in fact more economical than Subj raising under the MWC. The MCC, on the other hand, chooses Subj raising over Obj raising because only the former allows the Case feature checking of T. To the extent that Subj raising is obligatory in this language, the conclusion seems to be that the MCC outranks the MWC.<sup>7</sup>

Under this conception, the MWC comes into play only when comparing two operations of Attract which will check the same set of the target's features. In other words, the MWC prohibits the target from attracting a feature which will not check any of its features. After all, this is the essence of Greed and Last Resort (Chomsky 1993) as they are reinterpreted and incorporated into the definition of Attract-F in Chomsky (1995). In section

6, I will return to a case where the suggested ranking relation between the MCC and the MWC is apparently at stake.

#### 4.5 A Brief Summary

We have arrived at the following general picture of our theory of  $CH_L$ :

(41) Generalized Attract:

Attract is the only kind of operation in  $CH_L$ .

(42) Optimal Derivation:

An optimal derivation takes the most economical step at every stage in terms of local economy principles.

(43) Principles of Local Economy:

MLC: Attracting a closer element is more economical.

MCC: Attracting an element that checks more formal features of the target is more economical.

MWC: Attracting fewer formal features is more economical.

(44) Rankings:

The MLC and the MCC are negatively tied. ( $MLC =_{Neg} MCC$ )

The MCC outranks the MWC. ( $MCC \gg MWC$ )

I have not considered the relative status of the MLC and the MWC, but a natural speculation from (44a,b) is that the MLC also outranks the MWC. In the absence of evidence to the contrary, I assume that this is so.<sup>8</sup>

#### 5 Movement into a $\theta$ -Position

Let us extend our generalized Attract approach to other examples discussed by Chomsky (1994, 1995), to show how it blocks movement into a  $\theta$ -position. Consider (45a,b).

- (45) a. [TP<sub>1</sub> John T<sub>1</sub> [<sub>v</sub> P t<sub>j</sub> Vb=believes [TP<sub>2</sub> Mary T<sub>2</sub> to be t<sub>M</sub> clever ]]].  
 b. \* [TP<sub>1</sub> John T<sub>1</sub> [<sub>v</sub> P t'<sub>j</sub> Vb=believes [TP<sub>2</sub> t<sub>j</sub> T<sub>2</sub> to be Mary clever ]]].

In (45a), T<sub>2</sub> attracts *Mary* to check its EPP-feature, while *John* is directly attracted by the matrix verbal complex Vb = [<sub>v</sub> V *v*] for  $\theta$ -marking. In (45b), T<sub>2</sub> attracts *John*, which is then attracted by Vb for  $\theta$ -marking. *John* further raises to [Spec,TP<sub>1</sub>] in both derivations. Under Chomsky's (1994, 1995) analysis, despite the fact that (45a) has employed the more costly operation of Move in the lower clause, it can be chosen as optimal because (45b) does not converge. In particular, it is claimed that  $\theta$ -marking properties do not trigger Move so that movement into [Spec,<sub>v</sub> P] in (45b) violates Greed of Chomsky (1993, 1994), now incorporated into the definition of Move. *John* does not receive any  $\theta$ -role, nor *v* satisfies its  $\theta$ -grid, and the derivation crashes because of a  $\theta$ -Criterion violation.

In our approach, (45a,b) call for a rather different explanation. Notice first that since Merge is now as costly as Move, attracting *Mary* in the lower clause of (45a) should pose no puzzle; it is as economical as attracting *John* in terms of the MWC as well as the MLC and the MCC. The real problem is rather that our approach permits movement of *John* to [Spec,<sub>v</sub> P] in (45b) per se. This should be so, to the extent that satisfaction of  $\theta$ -marking properties is required by convergence.

In addressing a similar issue, Collins (1995) departs from Chomsky (1994) and proposes to exclude a derivation as in (45b) as nonconvergent on grounds that the covert Case checking of *Mary* violates the MLC. I adopt this line of analysis for our account of (45a,b), which will lead us to abandon Chomsky's (1995) claim that traces are immobile.

- (46) Only the head of a chain enters into the operation Attract/Move.

(Chomsky 1995; 304)

In our analysis, the MLC is retained as an economy principle so that its violation does not necessarily yield nonconvergence, contra Chomsky (1994) and Collins (1994). As I will show, derivation (45b) does converge, but not optimally.

I also claim, contra Chomsky (1995), that the covert accusative Case checking takes place by feature adjunction to Vb before it raises to T.

- (47) a. [TP Subj T ... [<sub>v</sub>P t<sub>Subj</sub> Vb [VP tv Obj ]]]  
 b. [TP Subj T ... [<sub>v</sub>P t<sub>Subj</sub> FF(Obj)-Vb [VP tv t<sub>FF(Obj)</sub> ]]]  
 c. [TP Subj [T FF(Obj)-Vb-T][<sub>v</sub>P t<sub>Subj</sub> tvb [VP tv t<sub>FF(Obj)</sub> ]]]

Suppose the language does not allow overt object shift. The relevant structure before Spell-Out is (47a). I propose that, after Spell-Out, FF(Obj) raise to adjoin directly to Vb as in (47b), rather than to T after Vb has raised to T as in (47c). (47b) differs from (47c) in that it does not allow FF(Obj) to c-command t<sub>Subj</sub>, which is presumably a desirable result.<sup>9</sup>

The attraction of FF(Obj) in (47b) satisfies the MLC, but its legitimacy in (47c) depends on how we interpret the status of the intervening t<sub>Subj</sub>. Chomsky (1995), in arguing for a derivation as in (47c), proposes to avoid a potential MLC violation here by assuming (46) above. Thus FF(Obj) can skip t<sub>Subj</sub> and adjoin to T without violating the MLC. If traces are generally immobile in this way, however, it is hard to see how derivation (45b) can be blocked by the MLC in our analysis. I therefore abandon (46) and propose (48).

- (48) Trace has exactly the same feature constitution as the head of its chain.



(48) is an immediate consequence of Chomsky's (1993) copy theory of movement, according to which movement of  $\alpha$  leaves behind a strict copy of  $\alpha = \text{trace of } \alpha$ . Suppose that in a chain  $\text{CH}=(\alpha, t)$ ,  $\alpha$  has feature F. It follows from (48) that not only  $\alpha$  but  $t$  can be attracted by the target  $\tau$  with F. Likewise,  $t$  rather than  $\alpha$  may attract an element for feature checking. As a result, covert FF(Obj) raising may target the trace of Vb, even in languages with overt V-raising, yielding an LF structure which is the same as (47b) in every relevant respect, thereby satisfying the maximum uniformity of LF outputs (see footnote 9).

This much in mind, let us return to (45a,b) and consider how their derivations may proceed after Spell-Out. (45a) is followed by the covert raising of FF(*Mary*) to Vb, as shown in (49).

(49) ... [<sub>vP</sub> t<sub>J</sub> FF(M)-Vb [<sub>TP2</sub> t<sub>FF(M)</sub> T2 to be t<sub>M</sub> clever ]]

There is no MLC violation here, since Vb attracts the closest element that may check its formal features. The derivation converges in the optimal way. On the other hand, (45b) may proceed as in (50).

(50) ... [<sub>vP</sub> t'<sub>J</sub> FF(M)-Vb [<sub>TP2</sub> t<sub>J</sub> T2 to be t<sub>FF(M)</sub> clever ]]

This step violates the MLC because of the intervening t<sub>J</sub> or its formal features FF(t<sub>J</sub>). Note that *John*, and therefore its trace under (48), have no Case feature but still retain the [+Interpretable] D- and  $\phi$ -features, whereas Vb has all these features unchecked at this stage of the derivation. The MLC requires then that Vb attract FF(t<sub>J</sub>) instead of FF(M). Suppose it does, as in (51).

(51) ... [<sub>vP</sub> t'<sub>J</sub> FF(t<sub>J</sub>)-Vb [<sub>TP2</sub> t<sub>FF(t<sub>J</sub>)</sub> T2 to be Mary clever ]]

Comparing (50) and (51), the two possible continuations of (45b), we see that neither step can be optimal; (50) violates the MLC, but (51) violates the MCC instead, because it fails to check the Case feature of Vb. This is a situation already familiar from our account of (28a-c) in the previous section, where the MLC and the MCC conspire to block (28b,c) simultaneously. In exactly the same manner, (50) and (51) are excluded together, deadlocking the derivation of (45b), and (45a) is correctly chosen as an optimal derivation.

We now successfully maintain the maximal simplicity and generality of the operation of Attract; movement into a  $\theta$ -position need not be precluded as such, but a derivation involving such an operation can be blocked as nonoptimal.

## 6 Optimal Derivations of Transitive Clauses

In this section, I will show that our approach explains in a unified manner why transitive clauses never seem to allow the pair of accusative Subj and nominative Obj, without recourse to certain auxiliary assumptions Chomsky (1995) adopts to derive the same effect in his approach, including the notion of derivation cancellation under feature mismatch (see (27) above). The general story goes as follows: irrespective of the actual feature constitutions of  $v$  and T involved, a derivation containing such a Subj-Obj pair never converges in an optimal way as it always violates either the MLC or the MCC.

### 6.1 Weak $v$ and Strong T

Suppose first that  $v$  has a weak D-feature but T has a strong D-feature, so that there is no overt object shift, but [Spec,TP] must be filled overtly. The two relevant structures, one containing Subj<sub>[nom]</sub> and Obj<sub>[acc]</sub> and the other Subj<sub>[acc]</sub> and Obj<sub>[nom]</sub>, are (52a,b), respectively.

- (52) a. [TP T [<sub>v</sub>P Subj<sub>[nom]</sub> *v* [VP V Obj<sub>[acc]</sub>]]]  
 b. [TP T [<sub>v</sub>P Subj<sub>[acc]</sub> *v* [VP V Obj<sub>[nom]</sub>]]]

What is to be shown is that (52a) has an optimal convergent derivation but (52b) has none. Importantly, we are not comparing (52a) and (52b) because they do not share the same initial numeration.

Consider first (52a), which has the following two possible continuations, according to which of the two DPs is attracted by T:

- (53) a. [TP Subj<sub>[nom]</sub> T [<sub>v</sub>P tSubj *v* [VP V Obj<sub>[acc]</sub>]]]  
 b. [TP Subj T [<sub>v</sub>P tSubj FF(Obj)-*v* [VP V tFF(Obj)]]]  
 (54) a. [TP Obj<sub>[acc]</sub> T [<sub>v</sub>P Subj<sub>[nom]</sub> *v* [VP V tObj]]]

(53) represents a convergent derivation, in which T attracts Subj pre-Spell-Out and *v* (or its trace; see the discussion in the previous section) attracts FF(Obj) post-Spell-Out. Turning to (54), suppose, tentatively following Chomsky (1995), that it is a canceled derivation under feature mismatch (between the nominative feature of T and the accusative feature of Obj). Such cancellation amounts to convergence in economy comparisons so that it cannot be avoided by choosing a more costly step. One can safely describe this state of affairs in the form of the following requirement on cancellation:

- (55) Cancel the derivation unless it converges more economically.

Thus, if (54a) were more economical than (53a), it would incorrectly block the convergent derivation (53) and the Subj<sub>[nom]</sub>-Obj<sub>[acc]</sub> pair would never surface in a transitive construction.

As a matter of fact, step (54a) is more costly than (53a) under the MLC, so that this wrong prediction does not come out. Rather, the apparent puzzle

is how (53) counts as optimal. To the extent that we adopt the global SDC (12), it is hard to solve this puzzle because convergence of (53) requires more steps than cancellation of (54). With the local MCC replacing the SDC, however, we compare only (53a) and (54a), to find that (53a) is more economical also with respect to the MCC, because the nominative Case feature of T is checked only by this step. In this way, step (54a) is excluded by both the MLC and the MCC in a strictly local manner, yielding the correct prediction that  $\text{Obj}_{[\text{acc}]}$  never appears in  $[\text{Spec}, \text{TP}]$ . Cancellation under feature mismatch does not play a part in this account.

Consider now (52b), whose possible continuations are as in (56) and (57):

(56) a.  $[\text{TP } \text{Subj}_{[\text{acc}]} \text{ T } [_\nu \text{P } \text{tSubj } \nu \text{ [VP } \text{V } \text{Obj}_{[\text{nom}]}]]]$

(57) a.  $[\text{TP } \text{Obj}_{[\text{nom}]} \text{ T } [_\nu \text{P } \text{Subj}_{[\text{acc}]} \nu \text{ [VP } \text{V } \text{tObj} ]]]$

Chomsky (1995) would take (56) as a canceled derivation. Again, our analysis dispenses with such an artifact and excludes both (56a) and (57a) at the same time. Namely, the MLC rejects (57a) in favor of (56a), but the MCC forces the opposite choice; attracting  $\text{Subj}_{[\text{acc}]}$  as in (56a) fails to check the nominative Case feature of T. Thus neither step is optimal, and the derivation is terminated at this stage.

## 6.2 Strong $\nu$ and Strong T

Suppose next that both  $\nu$  and T have strong D-features, so that both overt object shift and the EPP effect are mandatory. We will see that only (58a) has an optimal derivation, in which Obj first raises to the outer  $[\text{Spec}, \nu \text{P}]$  and then Subj raises to  $[\text{Spec}, \text{TP}]$ , with all other possible derivations being blocked.

- (58) a. [<sub>v</sub> P Subj<sub>[nom]</sub> *v* [VP V Obj<sub>[acc]</sub>]]  
 b. [<sub>v</sub> P Subj<sub>[acc]</sub> *v* [VP V Obj<sub>[nom]</sub>]]

Given in (59-61) are the three possible continuations of (58a), of which (59) must be the optimal derivation:

- (59) a. [<sub>v</sub> P Obj<sub>[acc]</sub> [<sub>v</sub>' Subj<sub>[nom]</sub> *v* [VP V t<sub>Obj</sub> ]]]  
 b. [TP Subj<sub>[nom]</sub> T [<sub>v</sub> P Obj [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V t<sub>Obj</sub> ]]]]  
 (60) a. [<sub>v</sub> P Obj<sub>[acc]</sub> [<sub>v</sub>' Subj<sub>[nom]</sub> *v* [VP V t<sub>Obj</sub> ]]]  
 b. [TP Obj T [<sub>v</sub> P t'<sub>Obj</sub> [<sub>v</sub>' Subj<sub>[nom]</sub> *v* [VP V t<sub>Obj</sub> ]]]]  
 c. [TP Obj FF(Subj)-T [<sub>v</sub> P t'<sub>Obj</sub> [<sub>v</sub>' t<sub>FF(Subj)</sub> *v* [VP V t<sub>Obj</sub> ]]]]  
 (61) a. [<sub>v</sub> P Subj<sub>[nom]</sub> [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V Obj<sub>[acc]</sub>]]]

(59) and (60) share step (i), in which *v* attracts Obj<sub>[acc]</sub> to its outer Spec, whereas in (61a) it is Subj<sub>[nom]</sub> that is attracted first. In (59b), T then attracts Subj<sub>[nom]</sub> to [Spec,TP] so that the derivation converges immediately. In (60b), however, T attracts the Case-less Obj instead so that convergence requires one more application of Attract after Spell-Out to check the still unchecked Case feature of T, as in (60c).

Comparing only these two derivations, Chomsky (1995) does propose that (60) is blocked by (59) because of this additional step – it violates the SDC. In this light, consider (61a), in which *v* attracts Subj<sub>[nom]</sub> to its outer Spec; this would be an instance of cancellation in Chomsky's approach, which raises the serious problem of why this canceled derivation does not count as optimal under the SDC. Cancellation of (61) involves fewer steps than convergence of (59) or (60), and it is most economical under the SDC. It is now obvious why we should reject the global SDC, together with the notion of cancellation under feature mismatch.

In our analysis, the explanation is again strictly local. We first compare steps (59a/60a) and (61a), choosing the former in terms of the MCC; there is no need to cancel (61) under feature mismatch. Next we compare (59b) and (60b), choosing the former again in terms of the MCC. Notice that the MWC would favor the latter instead, since the Case-less Obj is 'lighter' than the nominative Obj; given our earlier conclusion that the MCC outranks the MWC, only (59b) counts as an optimal step. Thus (59) is correctly picked out as an optimal derivation, as a result of two consecutive local economy comparisons in terms of the MCC.

Consider now the possible continuations of (58b), shown in (62-64):

- (62) a. [<sub>v</sub> P Subj]<sub>[acc]</sub> [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V Obj]<sub>[nom]</sub>]]]  
       b. [TP Obj]<sub>[nom]</sub> T [<sub>v</sub> P Subj [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V tObj ]]]]  
 (63) a. [<sub>v</sub> P Subj]<sub>[acc]</sub> [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V Obj]<sub>[nom]</sub>]]]  
       b. [TP Subj T [<sub>v</sub> P t'<sub>Subj</sub> [<sub>v</sub>' t<sub>Subj</sub> *v* [VP V Obj]<sub>[nom]</sub>]]]]]  
 (64) a. [<sub>v</sub> P Obj]<sub>[nom]</sub> [<sub>v</sub>' Subj]<sub>[acc]</sub> *v* [VP V tObj ]]]]

We want to show that none of these leads to an optimally convergent derivation. In (62a/63a), *v* first attracts Subj]<sub>[acc]</sub> to its outer Spec, whereas in (64a) it attracts Obj]<sub>[nom]</sub>, instead. The MCC rejects the latter step in favor of the former for the now familiar reason. Compare next (62b) and (63b), in which T attracts Obj]<sub>[nom]</sub> and the Case-less Subj to its Spec, respectively. Interestingly, (62b) is more economical under the MCC, but it is rather (63b) that is more economical under the MLC. The two principles now conspire to exclude both steps at once. The two consecutive local economy comparisons, first in terms of the MCC and next in terms of the MLC and the MCC, enable us to reject all of (62-64) straightforwardly. Notice again that there is no need to cancel (64) under feature mismatch.

Chomsky (1995; 294), in excluding the derivation as in (63) under his system, invokes the following economy principle on the determination of the initial numeration:

(65)  $\alpha$  enters the numeration only if it has an effect on output.

Recall that Chomsky reinterprets the MLC as part of the definition of Move, so that (62b) is not a possible step at all. The problem is therefore why (63) does not constitute an instance of the global SDC being overridden by convergence, which would render it an optimal derivation. Chomsky claims that the PF and LF outputs that will result from this derivation are the same as when the derivation does not involve the strong D-feature of  $v$  from the start. (65) prohibits selection of the strong  $v$ , and (63) can be safely ignored, so Chomsky concludes.

Notice, however, that (65) is a conceptually very odd principle, largely due to its radically global nature. How can one know whether or not a particular choice of a strong feature in the numeration affects an output, unless one can hypothetically let all the possible derivations proceed till they converge, both with and without the feature in question in the numeration, and compare the resulting outputs? Such hypothetical computations entail economy comparisons of extremely high complexity. Even worse, by necessitating comparisons among derivations with different numerations, (65) in fact multiplies the reference set virtually infinitely, leading to the kind of “exponential blowup” (Chomsky 1995; 228) that we are eager to avoid.

Furthermore, the observation that the choice of strong  $v$  in (63) does not have any effect on output is not correct, at least to the extent that PF is concerned. Recall that (63) is a derivation of the transitive construction with

Subj<sub>[acc]</sub> and Obj<sub>[nom]</sub>. Without the strong D-feature of *v*, such a derivation is inevitably canceled under Chomsky's system, but selecting this feature will have the effect of (incorrectly) allowing the derivation to converge. In short, selecting strong *v* in (63) is the only way of constructing a convergent derivation which will yield a PF output that would not be available otherwise. This means that (65) in fact does not block the selection of strong *v*, and that (63) cannot be excluded even by this highly dubious principle under Chomsky's system.

By contrast, our approach allows a very simple computation: economy comparisons are always strictly local, and Chomsky's (28) and (65) are both dispensable.

### 6.3 A Note on Inverse Voice

In our account of the optimality of (59) above, it has been argued that step (59b) blocks (60b) under the MCC, although the latter step is in fact more economical under the MWC. The proposed ranking relation between these two principles is crucial in maintaining this account.

There is an attested linguistic phenomenon that may appear to challenge this view. Ura (1996) examines Inverse voice in Bantu and proposes to accommodate it by allowing both of the two derivations (59) and (60) to converge optimally. Ura gives the following examples from Kinyarwanda.

(66) a. Active;

Umuhuungu	a-ra-som-a	igitabo
boy	he-Pres-read-Asp	book
'The boy is reading the book.'		

b. Inverse;



Igitabo cyi-ra-som-a umuhuŋgu

book it-Pres-read-Asp boy

'The book is being read by the boy.'

Ura argues that, given overt V-raising to T in this language, a derivation as in (59) yields the active SVO order in (66a), whereas a derivation as in (60) yields the inverse OVS order in (66b). Ura's analysis, if correct, constitutes counterevidence to our account of (59). On a closer examination, however, several possible solutions turn out available. Here I will sketch some of them.

In our account, the assumption that the MCC outranks the MWC allows us to choose (59) over (60). Suppose, then, that this ranking relation is somehow altered in the language under consideration. One possibility is that the MCC and the MWC are positively tied, so that steps (59b) and (60b) may count as equally economical, permitting both derivations to converge optimally. Another possibility is that the MWC outranks the MCC in this language, with the auxiliary assumption that PF convergence does not require pied-piping of the Case feature of a DP. Then even when T attracts Subj as in (59b), its nominative Case feature is stranded in situ, so that the operation may satisfy the MWC rather than the MCC. Then attracting the Case-less Obj to [Spec,TP] is exactly as economical as attracting Subj in the suggested manner, so that the two derivations again converge optimally.

Such OT-style accounts of language variation in terms of different rankings among economy principles are, however, obviously in contradiction to the fundamental Minimalist claim that  $C_{HL}$  is uniform and variation should be limited to the lexicon, in particular to its functional elements. I therefore suggest here that an even more plausible approach to the free voice

alternation is rather by noting the different manifestations of subject agreement in (66a,b). Notice that Subj induces subject agreement in the active (66a), whereas Obj does so in the inverse (66b). This contrast can be understood to indicate that the two derivations differ in their initial numerations, in particular with respect to the  $\phi$ -feature specification of T. If so, these derivations do not compete from the start. When T has Subj  $\phi$ -features, attracting Subj to [Spec,TP] is more economical than attracting Obj under the MCC, and the opposite is the case when T has Obj  $\phi$ -features, instead. Attracting Obj when T has Subj  $\phi$ -features, for example, is more costly because it fails to check the  $\phi$ -features of T, no matter how economical it is under the MWC. Accordingly, the relative status of the MCC and the MWC remains intact; the MCC outranks the MWC even in languages exhibiting Inverse voice.

## 7 Violability of the MLC

I have retained the MLC as an economy principle throughout, contra Chomsky (1995). This view entails that the MLC can be violated where the derivation would not converge otherwise, in accordance with the Minimalist claim that convergence overrides economy. In this section, I will support my approach by presenting and solving a problematic case which receives a natural account only if we take the MLC as a violable principle. The analysis to be proposed also involves a reformulation of Chomsky's (1986) Uniformity Condition as a condition on inherent Case checking.

Chomsky (1995) notes that the MLC as part of the definition of Attract/Move incorrectly rules out a sentence like (67a).

- (67) a. [TP<sub>1</sub> John T<sub>1</sub> seems to Mary [TP<sub>2</sub> t<sub>j</sub> to be intelligent ]].

- b. \*[TP<sub>1</sub> Mary T<sub>1</sub> seems to t<sub>M</sub> [TP<sub>2</sub> John to be intelligent ]].

Given that *Mary* is in a position c-commanding into TP<sub>2</sub>, the MLC requires that T<sub>1</sub> attract *Mary* rather than *John* to [Spec,TP<sub>1</sub>], predicting (67b) to be grammatical, contrary to fact. If the MLC is a violable economy principle, however, we can save (67a) by showing that the alleged competing derivation (67b) crashes, since (67a) will then be an instance of MLC violation forced by convergence. Let us see how this can be so.

I assume, following Chomsky (1995), that here *Mary* is  $\theta$ -marked and inherently Case-marked by the verb *seems*. Given that *John* has the nominative Case feature, it may appear that the familiar conspiring relation of the MLC and the MCC excludes both of (67a,b), since attracting *Mary* to [Spec,TP] fails to check the Case feature of T and is therefore more costly under the MCC. I propose that this is indeed the reason why the French counterparts of (67a,b) are both ungrammatical (see Chomsky 1995).

- (68) a. \*[TP<sub>1</sub> Jean semble-T<sub>1</sub> à Marie [TP<sub>2</sub> t<sub>J</sub> avoir du talent ]].  
       cf. \*Jean semble à Marie avoir du talent.  
       b. \*[TP<sub>1</sub> Marie semble-T<sub>1</sub> à t<sub>M</sub> [TP<sub>2</sub> Jean avoir du talent ]].  
       cf. \*Marie semble à Jean avoir du talent.

The problem is therefore why (67a) is not excluded in the same way. Given that only convergent derivations compete, it is naturally expected that while the French (68b) converges and blocks (68a), the English (67b) in fact crashes and therefore does not block (67a).

I propose that this English-French asymmetry derives from the commonly acknowledged parametric variation that overt V-raising applies in French but not in English (Pollock 1989, Chomsky 1993), in tandem with

a reformulation of Chomsky's (1986) Uniformity Condition.

(69) Uniformity Condition (UC):

If  $\alpha$  is an inherent Case-marker, then  $\alpha$  Case-marks NP iff

$[\alpha]$   $\theta$ -marks the chain headed by NP. (Chomsky 1986; 194)

The UC was introduced in part to rule out instances of illicit NP-internal raising, including (70).

(70) \* $[\text{NP John's belief } [\text{IP } t_j \text{ to be intelligent } ]]$

Here the inherent Case marker *belief* does not  $\theta$ -mark *John*. Violation of the UC counts as a  $\theta$ -Criterion violation, the latter of which I tentatively assume with Chomsky (1994, 1995) to cause a derivation to crash.

Roughly put, the UC requires that an inherent Case be licensed under strict locality between the assigner and the assignee. Let us assume, as seems natural, that this local relation is structurally captured by the Spec-head relation or other checking configurations available in the theory. I now propose to reformulate the UC as a condition on inherent Case checking:

(71) Revised Uniformity Condition (RUC):

If  $\alpha$  is an inherent Case-marker, then  $\alpha$  Case-checks  $\beta$  iff

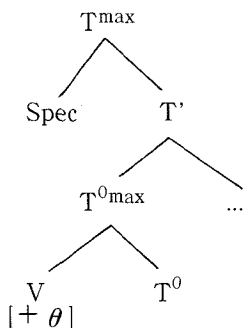
(i)  $\alpha$   $\theta$ -marks  $\beta$ , and (ii) the Case feature and  $\theta$ -feature of  $\alpha$  bear the same structural relation to those of  $\beta$ .

Here the term ' $\theta$ -feature' is generalized to refer to whatever is responsible for the  $\theta$ -marking relation in question. For  $\alpha = V$ , for example, it corresponds to its  $\theta$ -marking property or its  $\theta$ -grid; for  $\beta =$  complement of  $\alpha$ , it corresponds to its  $\theta$ -role assigned by  $\alpha$ , and so on. Crucially,  $\theta$ -features in this sense are semantic features and are inaccessible to the computational system.

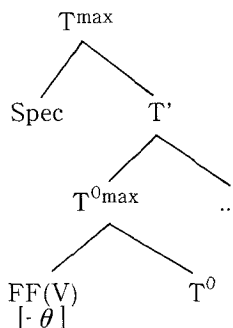
The ungrammaticality of (70) follows from the RUC as before; *John* is in a checking configuration with *belief*, but the Case checking is impossible because *belief* does not  $\theta$ -mark *John*. The inherent Case remains unchecked and this causes the derivation to crash, whether or not an RUC violation counts as a  $\theta$ -Criterion violation, too.

The RUC provides an illuminating account of why (67b) crashes but (68b) converges, leading us to the correct explanation of the English-French asymmetry above. Consider French first, where V (or the V-*v* complex, more precisely) raises to T before Spell-Out and yields the structure (72a):

(72) a. French



b. English



[Spec,TP] in this language counts as a checking position of an inherent Case assigned by V, because the overt V-raising takes the form of a category movement, pied-piping whatever features constitute V, including its  $\theta$ -feature. In (68b), then, *semble* Case-checks *Marie* in [Spec,TP1] under the RUC, in particular because their Case features and  $\theta$ -features share the same Spec-head relation, and the derivation converges as desired. Note that the Case feature of T1 is checked not by *Marie* but by the covertly raised FF(*Jean*), exactly as in the expletive construction. Since both of (68a,b)

converge, they compete with and block each other, under the conspiracy of the MLC and the MCC.<sup>10</sup>

In English, V never raises to T before Spell-Out. More importantly, even after Spell-Out, what raises to T is FF(V), excluding its  $\theta$ -feature, as in (72b). [Spec,TP] therefore never counts as a checking position of an inherent Case assigned by V even at LF, where the RUC comes into play after all. In (67b), *seems* fails to Case-check *Mary* in [Spec,TP1] because their  $\theta$ -features do not share the same structural relation as their Case features. (67b) crashes, and therefore it does not block (67a) under the MLC. In short, the MLC violation in (67a) is a good example of an economy principle being overridden by convergence. This consideration strongly supports our view that the MLC is an economy principle, contra Chomsky (1995).

## 8 Concluding Remarks

In this paper, I have attempted to further revise and improve the recent development of the Minimalist Program advanced by Chomsky (1995) along the following lines:

- (73) a. Attract is the only derivational operation in  $C_{HL}$ .
- b. Determination of an optimal derivation takes place in a strictly local manner under the interaction of the MLC, the MCC, and the MWC.

The generalized theory of Attract strengthens the uniformity of  $C_{HL}$  from N to  $\lambda$ . While many problems are left open for future research, I firmly believe that the approach just outlined deserves serious considerations for the purpose of further promoting the study of the human language faculty along the general guidelines of the Minimalist Program.

## Notes

\*The material presented here grew out of a series of talks I gave on various occasions including: Osaka Minimalism Circle (January 1995, at Osaka University), Sophia University Linguistic Colloquium (June 1995), Kansai Association for Theoretical Linguistics (October 1995, at Kobe University), the 13th National Conference of the English Linguistic Society of Japan (November 1995, at Tokyo Gakugei University), and Tohoku University English Linguistic Circle (February 1996). I am grateful to the participants for their comments, questions, and whatever made my presentations worth while. Special thanks go to Masaru Nakamura and Hiroyuki Ura. Usual disclaimers apply.

1. In fact, Chomsky (1994, 1995) maintains that a  $\theta$ -Criterion violation causes the derivation to crash, just as a [-Interpretable] formal feature that remains unchecked at  $\lambda$  does. A related question one might ask is: Can Move be motivated by  $\theta$ -marking? In other words, is there movement into a  $\theta$ -position? Chomsky (1994, 1995) definitely denies this possibility, distinguishing  $\theta$ -roles from formal features that require checking. I will return to this topic in section 5, where I will suggest that such movement is blocked for an economy reason.
2. (7a-c) are adapted from Chomsky's original examples for the sake of simplicity.
3. Chomsky (1994) assumes that (7c) is also nonconvergent even though all the relevant features are checked, stating that an MLC violation causes the derivation to crash. I reject this view and take both (7b) and (7c) to be convergent derivations. See the discussion immediately below.
4. As another instance of Merge for feature checking, Chomsky (1995; 311) discusses merger of *whether* and *if* in an embedded interrogative clause, as in:
  - (i) a. I wonder [<sub>CP</sub> whether Q [<sub>TP</sub> he left ]].
  - b. I wonder [<sub>CP</sub> [<sub>Q</sub> if Q ] [<sub>TP</sub> he left ]].
5. In fact, Chomsky (1995; 292) also considers the possibility that Procrastinate holds of Merge as well, in connection with the plausibility of covert merger.
6. In (29-31) and the following structural representations, I omit the effect of overt/covert verb raising and other irrelevant details. See the discussion in section 5.
7. The following examples, pointed out by Hiro Ura (personal communication), seemingly contradict this conclusion.
  - (ii) a. [<sub>TP1</sub> It T1 seems [<sub>CP</sub> [<sub>TP2</sub> there T2 is likely [<sub>TP3</sub> *t*<sub>there</sub> T3 to be someone in the room ]]]].

- b. \* $[_{TP1} \text{ It } T1 \text{ seems } [_{CP} [_{TP2} t_{it} T2 \text{ is likely } [_{TP3} \text{ there } T3 \text{ to be someone in the room } ]]]]$ .

In (ia), T2 attracts *there* to its Spec, whereas in (ib), it attracts *it*, instead. The former step is more economical than the latter under the MWC but is more costly under the MCC. How, then, can (ia) be optimal, given that the MCC outranks the MWC?

One possibility one may pursue is that (ib) does not converge, perhaps because the D-feature of the expletive fails to be checked by covert N-feature raising from the associate (see the discussion in section 4.3). Descriptively, in the expletive construction the formal features of the associate, FF(A), must be first adjoined to T, whose Spec the expletive occupies, so that the D-feature of Exp may attract the N-feature of FF(A) (see Chomsky 1995: 364).

- (ii)  $[_{TP} \text{ Exp } [_{T'} [_{T} \text{ FF(A)-T } ] \dots ]]$

Here what attracts FF(A) to T is primarily the Case feature of T that is not checked by Exp; as a result, Exp and FF(A) enter into the Spec-head configuration. In (ib), however, FF(A) are attracted rather by the Case feature of T1, which is not checked by *it*. The resulting structure is roughly as follows:

- (iii)  $[_{TP1} \text{ It } \text{FF(A)-T1 } \dots [_{TP3} \text{ Exp } T3 \dots ]]$

Here Exp and FF(A) do not constitute a Spec-head configuration, and N-feature raising from FF(A) to Exp cannot apply, causing the derivation to crash. If so, (ia) presents an example of the MCC being overridden for convergence.

8. Relevant examples are hard to find. Consider (i):

- (i)  $[_{TP} T1 [_{vP} \text{ someone expects } [_{TP} \text{ there } T2 \text{ to be a riot someday } ]]]]$

Attracting *someone* to [Spec, TP1] is more economical than attracting *there* under the MLC, whereas the latter is more economical than the former under the MWC. This does not necessarily indicate that the MLC outranks the MWC, since the former step is also more economical under the MCC.

9. That is, only (47b) guarantees that the relative height of FF(Obj) and FF(Obj) at  $\lambda$  is kept uniform cross-linguistically, whether or not overt object shift applies in the language. In Fujita (1993, 1996), I have argued that the kind of backward binding typically manifested by nonvolitional causative predicates, including psych verbs, can be explained in terms of the LF reconstruction, on the assumption that Obj (or FF(Obj)) c-commands  $t_{\text{Subj}}$  in these constructions. This analysis is obviously incompatible with the structure (47c), since (47c) would freely allow backward binding in transitive clauses in general. See Fujita (1993, 1996) for details.



10. For (68a) and also (67a), I simply assume that the inherent Case checking takes place in situ.

## References

- Bobaljik, Jonathan David. 1995. In terms of merge: Copy and head movement. In *MIT Working Papers in Linguistics* 27, ed. Rob Pensalfini and Hiroyuki Ura, 41-64. MITWPL, MIT, Cambridge, Mass.
- Bowers, John. 1993. The syntax of predication. *Linguistic Inquiry* 24:591-656.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel Jay Keyser, 1-52. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1994. Bare phrase structure. *MIT occasional papers in linguistics* 5. MITWPL, MIT, Cambridge, Mass.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1996. Some observations on economy in generative grammar. Unpublished ms., MIT.
- Collins, Chris. 1995. Toward a theory of optimal derivations. In *MIT Working Papers in Linguistics* 27, ed. Rob Pensalfini and Hiroyuki Ura, 65-103. MITWPL, MIT, Cambridge, Mass.
- Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3:283-341.
- Fujita, Koji. 1993. Object movement and binding at LF. *Linguistic Inquiry* 24:381-388.
- Fujita, Koji. 1996. Double objects, causatives, and derivational economy. *Linguistic Inquiry* 27:146-173.
- Grimshaw, Jane. 1993. Minimal projection, heads & optimality. Technical Report #4, Center for Cognitive Science, Rutgers University, New Jersey.
- Lasnik, Howard. 1995. Case and expletives revisited: On Greed and other human failings. *Linguistic Inquiry* 26:615-633.
- Pesetsky, David. 1995. Optimality principles of sentence pronunciation. Paper presented at the Workshop on Optimality in Syntactic Theory, MIT.
- Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20:365-424.
- Poole, Geoffrey. 1995. Constraints on local economy. Paper presented at the Workshop on Optimality in Syntactic Theory, MIT.
- Ura, Hiroyuki. 1996. *Multiple feature-checking: A theory of grammatical function splitting*.

Doctoral dissertation, MIT. Distributed by MITWPL, MIT, Cambridge, Mass.

Watanabe, Akira. 1993. *AGR-based Case theory and its interaction with the A' system*.

Doctoral dissertation, MIT. Distributed by MITWPL, MIT, Cambridge, Mass.

Koji Fujita

Faculty of Integrated Human Studies

Kyoto University

Yoshida, Sakyo-ku

Kyoto 606-01

Japan

*kfujita@french.ic.h.kyoto-u.ac.jp*